

UNITED STATES PATENT APPLICATION

**TECHNIQUE FOR IMPROVING TRANSMISSION CONTROL PROTOCOL  
PERFORMANCE IN LOSSY NETWORKS**

**INVENTORS**

**Rohit Ramani**

**Abhay Karandikar**

Schwegman, Lundberg, Woessner & Kluth, P.A.  
1600 TCF Tower  
121 South Eighth Street  
Minneapolis, MN 55402  
ATTORNEY DOCKET SLWK 1488.011US1  
Client Ref. No. SCT/PF7

# TECHNIQUE FOR IMPROVING TRANSPORT PROTOCOL PERFORMANCE IN LOSSY NETWORKS

5

## Field of the Invention

This invention generally relates to transport protocol for communication networks having lossy links, and more particularly to a technique to improve the transport protocol performance in lossy networks including congestion control mechanisms.

10

## Background

Most Internet data applications today use reliable transport protocols, such as the transmission control protocol (TCP), which have been tuned for traditional networks comprising wired links and stationary hosts. TCPs assume congestion in the networks to be the primary cause for packet losses (packets can contain thousands of bytes of data) and unusual delays. Congestion occurs when the primary requirements of source(s) exceed the transport capability of a network node. For example, where multiple senders transmit packets to a network node faster than the node can forward the packets the buffer may overflow resulting in congestion, and some of the received packets can be lost. Congestion may also be defined as the condition that exists when the network node buffers occupancy is greater than some threshold.

The TCP protocol controls congestion by utilizing acknowledgements from the receiver and adjusting a sliding window for the sender. The sender will not be able to transmit packets beyond the window size. Upon receiving the acknowledgements from the receiver the window slides ahead and gets incremented. As long as there is no loss, the window size is gradually increased. When a loss occurs, the window size is reduced and then slowly expanded. The sender can identify that a packet has been lost due to congestion either by the arrival of duplicate acknowledgements indicating a loss or by the absence of receiving an acknowledgment within a timeout interval. This entire

process of controlling the window size to limit congestion is known as a “congestion control mechanism.”

When acknowledgements fail to be received by the sender for reasons other than congestion, however, congestion compensation measures, such as reducing the window size, can result in unnecessary reduction in end-to-end throughput and suboptimal performance. Today Internet and wireless communications links are increasingly being used to transmit data packets in applications such as email, web browsing, mobile computing, and other such applications. Most of these applications make use of reliable end-to-end transport level services provided by TCP. TCP was designed for networks where congestion is the primary cause of packet loss. However, transmission errors over wireless links such as due to fading may often lead to packet loss. Thus, the loss of packet may be due to reasons other than congestion. Wireless links often suffer from sporadic high bit-error rates (BERs) and intermittent connectivity problems due to handoffs. Consequently, TCP performance in networks having wireless links suffers from significant throughput degradation and very high interactive delays due to unnecessary use of congestion compensation mechanisms.

There are currently two approaches to improve TCP performance in such lossy links. A lossy link is a link where packet losses are due to high BERs rather than congestion, such as a wireless link. The first approach hides any non-congestion-related losses from the TCP sender, and therefore requires no changes to the existing sender implementation. The approach assumes that, since the problem is local, it should be solved locally, and that the transport layer need not be aware of the characteristics of the individual links. The lower layers of the network solve this problem by local retransmissions. Protocols adopting this approach attempt to make a lossy link appear to be of a higher quality at the cost of a reduced effective bandwidth. As a result, most of the losses seen by the sender are apparently caused due to congestion, thus reducing performance due to unwarranted invocation of the congestion control mechanism.

The second approach attempts to make the sender aware of the existence of lossy segments, by indicating to the sender that some packet losses are not due to congestion.

There are currently two methods using the second approach. The first method uses an end-to-end scheme named EE2E-ELN protocol. This scheme improves TCP performance in lossy networks by making the sender aware of the lossy link. The scheme adds Explicit Loss Notification (ELN) options to TCP acknowledgements. The problem with this scheme is that the Link Layer (LL) is involved in detecting packet losses. This is not desirable because it violates the modular structure of a TCP/IP suite. In addition, adding ELNs to TCP acknowledgements is not yet recommended by the Internet Engineering Task Force (IETF).

The second method uses a scheme, that uses the difference between Round Trip Time (RTT) of the last correctly received packet and the previously received packet, along with notification from the receiver about a packet loss, to determine the cause of the packet loss. The problem with this scheme is that it uses the difference between RTTs to find out reasons for a packet loss. For the scheme to be effective, RTT has to be precisely calculated. Because highly accurate clocks have to be used, it is impractical to implement this scheme in real-time applications. Therefore, there is a need in the art for a technique that improves TCP performance in lossy networks, such as wireless communication networks with congestion compensation mechanisms, that is computationally efficient and practical to implement.

### **Summary of the Invention**

The present invention provides an apparatus for an improved transport protocol for lossy networks. In one example embodiment, this is accomplished by using computationally efficient techniques to manage invoking congestion control mechanisms to control transmitted packets in lossy networks. The improved apparatus receives multiple packets including a header and an associated sequence number. The apparatus monitors the network for congestion caused by the received packets and marks the header of some of the packets with an impending congestion indication based on the outcome of the monitoring. The packets are then transmitted, and a receiver then sends back acknowledgements of receipt for each of the transmitted packets. Each of the

received acknowledgements includes an associated sequence number and any impending congestion indication assigned to the packet. The apparatus monitors each of the received acknowledgements, and invokes a congestion control mechanism based on the outcome of the monitoring to control a congestion window size that regulates the transmitted packets.

Another aspect of the present invention is a method for an improved transport protocol for lossy networks. The method is performed by receiving multiple packets including header and associated sequence number and monitoring the network for congestion caused by the received packets. The header of some of the packets is marked with an impending congestion indication, based on the outcome of the monitoring. The packets are then transmitted and acknowledgements of receipt are returned for each of the transmitted packets. Each of the received acknowledgements includes an associated sequence number and any impending congestion indication assigned to the packet. The method then monitors each of the received acknowledgements, and based on the outcome of the monitoring, invokes a congestion control mechanism to control a congestion window size that regulates the transmitted packets.

Another aspect of the present invention is a computer readable medium having computer-executable instructions for an improved transport protocol for lossy networks. According to the method, multiple packets including header and associated sequence number are received. The method then monitors the network for congestion caused by the received packets, and marks the header of some packet with an impending congestion indication, based on the outcome of the monitoring. The packets are then transmitted and acknowledgements of receipts are received for each of the transmitted packets. Each of the received acknowledgements includes an associated sequence number and any impending congestion indication assigned to the packet. The method then monitors each of the received acknowledgements, and based on the outcome of the monitoring, invokes a congestion control mechanism to control a congestion window size that regulates the transmitted packets.

Other aspects of the invention will be apparent on reading the following detailed description of the invention and viewing the drawings that form a part thereof.

### **Brief Description of the Drawings**

5           Figure 1 illustrates an embodiment of a communication system of the present invention.

          Figure 2 illustrates an example embodiment of a packet loss situation in a lossy network.

          Figure 3 is an example embodiment of a block diagram of a lossy network  
10       including major components of the TCP of the present invention.

          Figure 4 graphically compares the wireless topology of the prior-art TCP with the TCP of the present invention.

          Figure 5 is a flowchart illustrating the overall operation of the embodiment shown in Figure 3.

15           Figure 6 is a block diagram of a suitable computing system environment for implementing embodiments of the present invention, such as those shown in Figures 3 and 5.

### **Detailed Description**

          The present invention provides improved Transport Protocol such as  
20       Transmission Control Protocol (TCP) performance in lossy networks, such as wireless communication networks, which is computationally efficient and practical to implement.

          Figure 1 shows a communication system according to an embodiment of the present invention. Sender 110 is connected to a sender base station 115, which is  
25       connected to a communication network 120. The communication network 120 is connected to receive a base station 130. The receiver base station 130 is connected to receiver 140 over a lossy link 150. The lossy link 150 is a link where transmission losses are primarily due to interference, rather than congestion, such as a wireless link. Sender 110 and receiver 140 transmit and receive information through communication

network 120, respectively, through the receiver base station 130 and over the lossy link 150. The term 'information' includes data, text, voice, video, and other such information transmitted over a wireless link. It should be understood that this configuration is shown for simplicity of illustration and that the actual implementation can include many more communication devices, networks, and other such devices. To take advantage of the wireless link 150, receiver 140 can be a mobile receiver such as a laptop computer.

Figure 2 illustrates an example embodiment of a packet loss situation in a lossy network. Here the TCP source (*sender*) 210 is in the process of a transfer across a two-hop network 220 including a base station (*router*) 230 to a TCP receiver (*mobile host*) 240. At the depicted time, the line at congestion window 260 at the base station 230 includes 5 packets. Of the 5 packets transmitted 270 from the network 220, the first two packets are lost in the wireless link (*lossy link*) 280 due to wireless errors. Also shown in Figure 2, is the return of the acknowledgements 290 indicating receipt of the transmitted packets by TCP receiver 240. The acknowledgements include a consecutive sequence number associated with each of the transmitted packets. In the example shown in Figure 2, TCP receiver 240 has only received packets with sequence numbers 3, 4 and 5 because the first two packets are lost in the lossy link 280. Therefore, in this situation, TCP receiver 240 would return the acknowledgements with the sequence number 0, as it has already received the packet with the sequence number 0, indicating that the packet with sequence number 1 is lost and TCP source 210 should retransmit the packet with the sequence number 1. Only after TCP receiver 240 receives the retransmitted packet with the sequence number 1, it would send the acknowledgement back to TCP source 210 for the packet with the sequence number 1 and so on. It can be seen that TCP receiver 240 sends multiple acknowledgements associated with the sequence number of the last received packet until TCP receiver 240 receives the next transmitted packet.

Figure 3 illustrates one embodiment of a block diagram of a lossy network 300 including major components of a TCP of the present invention. The lossy network 300,

as shown in Figure 3, includes sender 110, which is coupled to sender base station 115, which in turn is coupled to the communication network 120. The lossy network 300, as shown in Figure 3, further includes receiver base station 130 coupled to communication network 130, which is coupled to the receiver 140. In this embodiment, communication  
5 network 120 includes an analyzer 320, a comparator 322, and a memory 324.

In operation, sender base station 115 receives multiple packets from a sender 110. Sender 110 can be a transmission control protocol source. Sender base station 115 then transmits the received packets through a lossy network. In this embodiment, the transmitted packets include a header and an associated sequence number. The header  
10 can include a congestion alleviation indication. The congestion alleviation indication can be in the form of marking a bit. In some embodiments, marking a bit includes setting to 1 to indicate that a step has been taken to reduce the congestion at the network caused by the transmitted packets. The header can also include an impending congestion indication. The impending congestion indication can be in the form of marking a bit,  
15 which can be set to 1 to indicate that the number of packets waiting in a line to be transmitted at sender base station 115 is in-between a predetermined minimum line size and a predetermined maximum line size.

Analyzer 320 then receives the transmitted packets from sender base station 115, and monitors each of the received packets for congestion and marks the header of some  
20 received packets with an impending congestion indication based on the outcome of the monitoring. In this embodiment, analyzer 320 monitors a number of received packets waiting in a line to be transmitted at sender base station 115. A comparator 322 then compares the number of received packets waiting in line at the sender base station 115 with predetermined minimum and maximum line sizes. Based on the outcome of  
25 comparison by comparator 322, analyzer 320 then marks the header of some of the received packets with the impending congestion indication. In some embodiments, analyzer 320 marks the header of some of the received packets when the number of packets waiting in line is greater than the predetermined minimum line size and less than the predetermined maximum line size based on a predetermined probability with an



impending congestion indication. In some embodiments, analyzer 320 drops the packets waiting in the line beyond the predetermined maximum line size when the number of packets waiting in line is greater than the predetermined maximum line size.

5 In some embodiments, analyzer 320 marks the header of some received packets by flagging bits in the header. Flagging bits in the header can include flagging ECN (Explicit Congestion Notification) flags like CE (Congestion Experienced) bits and CWR (Congestion Window Reduced) bit. Flagging CE bits can include setting CE bits to 1s. Setting CE bits to 1s indicates that the number of packets waiting in a line to be transmitted at sender base station 115 is in between a predetermined minimum line size  
10 and a predetermined maximum line size. Flagging a CWR bit can include setting the CWR bit to 1. Setting the CWR bit to 1 indicates that a step has been taken to reduce congestion at the network caused by the transmitted packets. In some embodiments, analyzer 320 provides a forward error correction to the header of each of the packets so that some of the packets that were lost due to wireless errors can be recovered back.

15 Receiver base station 130 then receives the transmitted packets from the communication network 120 and transmits the received packets through a lossy network 150. Receiver 140 then receives the transmitted packets from the lossy network and then returns acknowledgements to sender 110 through communication network 120 for each of the received packets through receiver base station 130. In this embodiment, the  
20 acknowledgements include a sequence number associated with each of the received packets and any associated marked impending congestion indication and congestion alleviation indication. In some embodiments, receiver 140 associates any marked impending congestion indication and congestion alleviation indication by flagging bits in acknowledgements of some of the received packets. Flagging the bits in the  
25 acknowledgements can include flagging ECN flags like ECE (ECN-Echo) bit. Flagging an ECE bit can include setting the ECE bit to 1. Setting the ECE bit to 1 indicates that the number of packets waiting in a line to be transmitted at the sender base station 115 is in between a predetermined minimum line size and a predetermined maximum line

size, and also that there is no indication that a step has been taken to reduce congestion at the network caused by the transmitted packets which means that flag CWR is set to 0.

Sender 110 then monitors each of the received acknowledgements for the sequence number and for any marked impending congestion indication. Based on the outcome of acknowledgement monitoring, sender 110 invokes a congestion control mechanism to control the transmission of packets to the sender base station 115.

In some embodiments, sender 110 monitors the received acknowledgements for a number of consecutively received acknowledgements with the same sequence number and then further compares the number of consecutively received acknowledgements with the same sequence number with a predetermined number and outputs a first signal when the number of consecutively received acknowledgements with the same sequence numbers is greater than or equal to the predetermined number. Upon receiving the first signal, sender 110 further checks the received acknowledgements for a marked impending congestion indication and invokes a congestion control mechanism when an acknowledgment is marked with a congestion impending indication.

In some embodiments, when a first signal is received and if the monitored acknowledgement is not marked with an impending congestion indication, then the congestion control mechanism is not invoked. In some embodiments, when a first signal is received and the monitored acknowledgement is not marked with an impending congestion indication, then sender 110 reduces the congestion window size by a predetermined small amount to regulate the transmitted packets.

Figure 4 is a graphical illustration 400 of a variation of TCP window size during a simulation time period of 0-60 seconds for both the prior-art TCP and the improved TCP according to the invention. The TCP congestion window size shown in Figure 4 is measured in terms of the number of packets in the congestion window. The number of packets in the congestion window is proportional to the number of packets transmitted by sender 110. Therefore, the areas under the dashed line graph 410 and solid line graph 420 are proportional to the number of packets transmitted by sender 110. Graphs 410 and 420 illustrate the number of packets in the congestion window at any given time in

the simulation time period when using the improved TCP according to the present invention and the prior-art TCP, respectively. It can be clearly seen that the area under the graph 410 is significantly larger than the area under the graph 420. From these results, it can be concluded that the improved TCP has a higher throughput than the prior-art TCP.

Graphs 410 and 420 were simulated using for a single sender and a single receiver and a network with a buffer capacity of 120 kbits, the data rate of the links set at 2 Mbps, and a propagation delay set at 2 microseconds. The 914 MHz Lucent Wave LAN radio interface was set using a power of transmission at 0.2818W, frequency at 914 MHz, and a receive threshold (the minimum power required to receive a packet at  $3.652 \times 10^{-10}$ ). For the ECN queue, the minimum line size  $\min_{th} = 96$  kbits, the maximum line size  $\max_{th} = 120$  kbits, mean packet size = 8 kbits, and the packet marking probability varies between 0 and  $\max_p$ , where  $\max_p = 0.02$ . The simulation illustrated in graphs 410 and 420 were performed for 60 seconds.

Figure 5 is a flowchart illustrating one example embodiment of a process 500 of providing a transmission control protocol within a lossy network. The process 500, as illustrated in Figure 5, controls a congestion window size, which regulates the transmitted packets. The process begins with step 510 by receiving multiple packets, including a header and an associated sequence number. The header can include a congestion alleviation indication. The header can also include an impending congestion indication. Step 520 includes monitoring the network for congestion caused by the received packets. In some embodiments, monitoring the network for congestion further includes monitoring the number of packets waiting in line to be transmitted. Then the number of packets waiting in line is then compared to a predetermined minimum line size and a predetermined maximum line size.

Step 530 includes marking the header of some of the packets with a congestion bit error, based on the outcome of the monitoring. In some embodiments, the header is marked with an impending congestion indication when the number of packets is greater than the predetermined minimum line size and less than the predetermined maximum

line size. In some embodiments, marking the impending congestion indication includes flagging CE bits in the header of some of the multiple packets based on a predetermined probability. In some embodiments, when the number of packets waiting in line is greater than the predetermined maximum line size then the packets waiting in line beyond the predetermined maximum line size will be dropped.

Step 540 includes transmitting the monitored packets through a lossy network. Step 550 includes returning acknowledgements after receiving each of the transmitted packets. Each of the transmitted acknowledgements includes a sequence number associated with each of the received packets and any associated impending congestion indication and the congestion alleviation indication. In some embodiments, an associated impending congestion indication and the congestion alleviation indication includes flagging an ECE bit in the acknowledgement of some of the transmitted packet.

Step 560 includes monitoring each of the received acknowledgements for the sequence number and the marked impending congestion indication associated with each of the received packets. In some embodiments, monitoring the received acknowledgements includes monitoring for a number of consecutively received acknowledgements with the same sequence numbers. Monitoring further includes checking the acknowledgements for a marked impending congestion indication when the number of consecutively received acknowledgements with the same sequence number is greater than or equal to a predetermined number. In some embodiments, the predetermined number of consecutive acknowledgements with the same sequence number is 3 or more.

Step 570 includes invoking a congestion control mechanism to control the transmitted packets based on monitoring the acknowledgements and the marked impending congestion indication. In some embodiments, the congestion control mechanism is invoked when the acknowledgment is marked with the impending congestion indication. In some embodiments, a congestion control mechanism is not invoked when the acknowledgement is not marked with an impending congestion indication. In some embodiments, when the acknowledgement is not marked with an

impending congestion indication then the congestion window size will be reduced by a predetermined small amount to regulate the transmitted packets.

Method 500, shown in Figure 5, may be implemented as a sender 110, a sender base station 110, a receiver base station 130, a receiver 140, and a communication  
5 network 120 including an analyzer 320, a comparator 322, and a memory 324 as shown in Figure 3.

Figure 6 shows an example of a suitable computing system environment 600 for implementing embodiments of the present invention, such as those shown in Figures 3 and 5. Various aspects of the present invention are implemented in software, which may  
10 be run in the environment shown in Figure 6 or any other suitable computing environment. The present invention is operable in a number of other general purpose or special purpose computing environments. Some computing environments are personal computers, server computers, hand-held devices, laptop devices, multiprocessors, microprocessors, set top boxes, programmable consumer electronics, network PCs,  
15 minicomputers, mainframe computers, distributed computing environments, and the like. The present invention may be implemented in part or in whole as computer-executable instructions, such as program modules that are executed by a computer. Generally, program modules include routines, programs, objects, components, data structures and the like to perform particular tasks or to implement particular abstract  
20 data types. In a distributed computing environment, program modules may be located in local or remote storage devices.

Figure 6 shows a general computing device in the form of a computer 610, which may include a processing unit 602, memory 604, removable storage 612, and non-removable storage 614. Memory 604 may include volatile memory 606 and non-  
25 volatile memory 608. Computer 610 may include – or have access to a computing environment that includes – a variety of computer-readable media, such as volatile memory 606 and non-volatile memory 608, removable storage 612 and non-removable storage 614. Computer storage includes RAM, ROM, EPROM & EEPROM, flash memory or other memory technologies, CD ROM, Digital Versatile Disks (DVD) or

other optical disk storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium capable of storing computer-readable instructions. Computer 610 may include or have access to a computing environment that includes input 616, output 618, and a communication connection 620.

- 5 The computer may operate in a networked environment using a communication connection to connect to one or more remote computers. The remote computer may include a personal computer, server, router, network PC, a peer device or other common network node, or the like. The communication connection may include a Local Area Network (LAN), a Wide Area Network (WAN) or other networks.

10

### **Conclusion**

- The above-described invention provides an improved Transport Protocol performance in lossy networks, such as wireless communication networks. The present invention accomplishes this by using computationally efficient techniques that are
- 15 practical to implement.

The above description is intended to be illustrative, and not restrictive. Many other embodiments will be apparent to those skilled in the art. The scope of the invention should therefore be determined by the appended claims, along with the full scope of equivalents to which such claims are entitled.